Model-based prosodic analysis of charismatic speech

Mixdorff, Hansjörg; Niebuhr, Oliver; Hönemann, Angelika

Published in:
Proc. 9th International Conference on Speech Prosody 2018

DOI:
DOI: 10.21437/SpeechProsody.2018-164

Publication date:
2018

9th International Conference on Speech Prosody 2018
13-16 June 2018, Poznań, Poland

Model-based prosodic analysis of charismatic speech

Hansjörg Mixdorff[1] *, Oliver Niebuhr*[2]*, Angelika Hönemann*[1]

[1]Beuth University Berlin, Germany
[2]SDU Electrical Engineering, Mads Clausen Institute, University of Southern Denmark
mixdorff@beuth-hochschule.de,oniebuhr@mci.sdu.dk,ahoenemann@beuth-hochschule.de

Abstract

This study examines at a new level of quantitative detail the intonation and timing properties of charismatic speech by comparing two popular CEOs, Steve Jobs and Mark Zuckerberg, who are known from informal observations and formal perception experiments alike to be more or less charismatic speakers, respectively. By applying the Fujisaki model we decomposed F0 contours into baseline frequency, phrasal F0 excursions and pitch accent-associated F0 excursions. Timing details are examined by applying Pfitzinger's model of perceived local speech rate to phone and syllable segmentations. Results suggest that high pitch not only involves generally higher F0 levels, but that these increases in F0 are not the same for every prosodic domain or level of the Fujisaki model. In addition we found significant differences depending on whether customers or investors are addressed.

Index Terms: Charisma, F0, Speech rate, Fujisaki model.

1. Introduction

Attracting attention as well as gaining and persuading followers without having any formal authority is the essence of charisma [7]. Or, in the words of [5]: "Charisma gets people to like you, trust you, and want to be led by you" (p.2). Charismatic speakers receive higher performance ratings and are viewed as more effective by both superiors and subordinates [5]. The subordinates also experience their work as more meaningful and make that little bit of extra effort above and beyond their call of duty when being led by a charismatic person [5]. Patients who perceive their physician to be charismatic are more likely to adhere to the given diagnosis and medical treatment and are less likely to file suit against him/her if things go wrong [6]. Moreover, the "wow factor" [6] charisma leads to more fruitful brainstorming outputs [1], results in better learning outcomes of students [2], helps entrepreneurs raise more start-up funding [3], and makes a product or service appear more credible and likable to customers [4]

Besides revealing and analyzing the positive effects of charisma for both the speakers and their social environment, previous studies also demonstrated that charisma is not a mysterious talent of a few gifted people, as was (and some-times still is) assumed by some sociologists [8,23], but a continuously varying skill that anyone can learn and improve [2,5,7,9]. In addition, these more learning-oriented studies have shown that for becoming a charismatic speaker "the non-verbal messages [of one's performance...] are more likely to influence the listeners' perceptions" [9:360] than the verbal messages, see also [24,25]. In other words, what we say is less important than *how* we say it, and here, the speaker's tone of voice - as prosody is typically called in rhetoric research and practice - turned out be play a crucial role [11,25]. [5:20] explains this crucial role by the fact that, "in the scope of human evolution, language is a relatively recent invention [whereas the age-old...] non-verbal communication is hard-wired into our brains". Therefore, unlike words, "it bypasses our logical thinking" (p.144). Simply put, without referring to them explicitly, [5] traces back the tremendous impact of non-verbal characteristics on speaker charisma to what phoneticians know as the "biological codes" [27,28].

In fact, quite a bit of phonetic research has been conducted in order to shed light on the links between prosodic features and charisma perception. With a focus on political speakers, a number of relevant (and to some extent culture-specific) prosody-charisma links have been worked out. They involve features like the level, range, and variability of F0 and intensity, the speaking rate, and the utterance-to-pause duration ratio. Previous studies also broke down charisma into simpler attributes with which the identified prosodic features are strongly correlated [11,13,14,15, 16,17,18,19]. Recently, our own studies showed that similar prosodic indicators of charisma apply to business speakers and political speakers alike. Moreover, our research added further parameters like the frequency of emphatic accents, voice quality (HNR, jitter, H1-H2), rhythmic variability (VarcoV, %V), and frequency of disfluency to list of charisma-relevant features [20] and addressed the issue of gender-specific behavior [26]. In addition, we extended the list of charisma-relevant features into the domain of sound segments, supporting the old claim of rhetoric that a "clear" and "crisp" articulation of "every phrase and word" [21:158] "is imperative to develop charisma" [22:138].

The present study continues our research on charisma in business speakers and takes a closer look at two of the probably most popular CEOs of our times: Steve Jobs (SJ) and Mark Zuckerberg (MZ). While both are or were successful company leaders and experienced public speakers, they strongly differ in the charismatic impression they make. SJ is celebrated as "a master of the art of effective and persuasive speaking" [40], whereas MZ's public speaking skills were described to be "rough enough to impact Facebook's perception in a negative way" [41]. These assessments from the media are consistent with findings from our own perception experiments. We took a randomly selected 30-second excerpt of a keynote speech from each of the two speakers. Both excerpts were in our ears prosodically representative of SJ's and MZ's public speech and contained no greater pauses or disfluencies. Then, we de-lexicalized the two 30-second excerpts by applying a low-pass filter to them (0-600 Hz) in PRAAT [29]. The resulting two stimuli were integrated in a set of 18 further de-lexicalized stimuli as part of a larger experiment (which we will report elsewhere). The stimuli were played in individually randomized orders to 98 participants whose task was to listen to the stimuli and answer three questions afterwards. One of them directly asked for speaker charisma: "How high would you rate the speakers charisma on a scale from 0-10"? The other two questions asked listeners to estimate the speaker's

 10.21437/SpeechProsody.2018-164

leadership/management experience (0-10 years) and the likelihood of investing their own money into the speaker's company (0-100%). As is shown in Figure 1, SJ clearly and statistically significantly outperformed MZ on the charisma question ($t[97]=11.5$, $p<0.001$) and the management/leadership question ($t[97]=24.8$, $p<0.001$). The investment question yielded no differences between the two speakers.

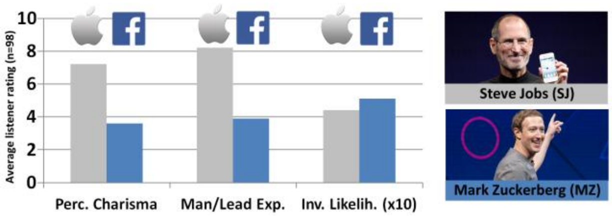

Figure 1: *SJ's and MZ's charisma* judged in terms of three questions by 98 listeners.

Based on our two speakers, we aim here to go beyond previous studies on prosodic charisma that were mainly concerned with holistic parameter settings and their differences from which it was, for example, concluded that a higher F0 level, a larger F0 range, a faster speaking rate, and more variability sound more charismatic [13-18]. We build on these general conclusions and use the Fujisaki model to look in more syntagmatic detail at what higher, larger, faster, and more variable actually mean at different positions and for different building blocks in the prosodic phrase. Moreover, we will differentiate between customer- and investor-oriented sections in SJ's and MZ's speech in order to address a further potentially relevant context factor in charisma production and perception. Note that we do not assume that everything SJ does is automatically more charismatic than what MZ does, since MZ is only less charismatic than SJ but not entirely uncharismatic. Yet, we expect SJ to outperform MZ in our acoustic analysis, especially in the critical F0 features.

2. Speech Material

We analyzed keynote speeches in the form of product presentations as these globally broadcasted events are particularly often referred to in the literature when it comes to speaker charisma [30]. For SJ, we used two of his most well-known and influential keynotes: the presentation of the iPhone 4 in 2010, and the presentation of the iPad 2 in 2011; 22 (10+12) minutes of speech were extracted from the two presentations. MZ's speech was extracted from two of his keynotes held in 2014 and 2015 at the annual F8 meeting of Facebook. MZ's keynote excerpts also comprise about 22 (11+11) minutes of speech. The two speech samples of MZ and SJ were extracted from the middle of their keynotes to exclude the usual opening and closing rituals of their presentations and potential biases due to the speakers' warm-up phases.

The structure the SJ's and MZ's is basically comparable: Showcasing the new product (new website features or functionalities in the case of Facebook) is followed by notes on the company's last year's achievements, position in the market, and future growth strategy. Obviously, the first showcasing section aims to motivate customers to buy or use the new product or service, whereas the following section on achievements, market and growth potential aims to keep or get more investors onboard. Since the speakers address two entirely different types of audience in the two sections of their keynotes (with entirely different intentions) and since audience type is known to be a relevant factor in charismatic speech [5,31], we will analyze the customer- and investor-oriented speech sections separately in our paper.

The audio files themselves were obtained from high-quality YouTube videos of the corresponding keynote presentations. Video and audio were separated, and the audio files were saved in the uncompressed WAV file format. For the segmental annotation, the speech samples of SJ and MZ were first orthographically transcribed by an English speaker and then submitted to automatic segmentation based on WebMaus and Darla [32,33]. For better handling we chunked the master wave files into pieces of about ten seconds containing complete inter-pause segments. Based on the word and phone transcriptions we manually added a syllable tier in *Praat TextGrid* [29].

3. Method of Analysis

We extracted F0 contours using *Praat*'s standard method [29] at equal steps of 10ms. This raw F0 contour is noisy and interrupted during unvoiced sounds. It is therefore difficult to relate directly to syllables of speech, for instance. This can be observed in Figures 2 and 3 which display the waveforms and F0 contours of two examples from the corpus.

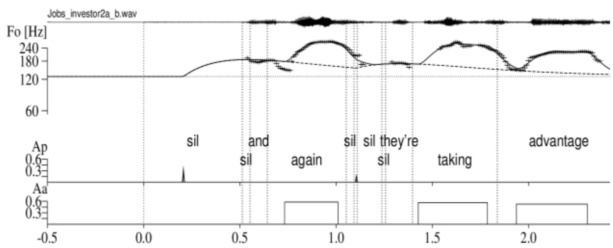

Figure 2: *Example of F0 decomposition of SJ's utterance "and again, they're taking advantage…"*

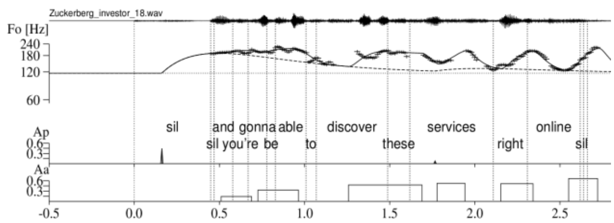

Figure 3: *Example of F0 decomposition of MZ's utterance "and you're gonna be able to discover these services right online …"*

In both panels the extracted *F0* (Hz) contour is indicated along the time axis by + signs underneath the waveform of the speech signal and the words of text, the boundaries of words are indicated by vertical dotted lines.

Since the sampled natural F0 contour does not lend itself well for direct comparative quantitative analysis, we employ the Fujisaki model [34] which produces a continuous *F0* contour from a parsimonious representation consisting of the superposition of three components: (1) the base frequency *Fb*, indicated by the horizontal line at the bottom of the *F0* pattern, (2) the phrase component, the slowly drooping phrasal contours accompanying each prosodic phrase, and (3) the accent component, reflecting fast *F0* movements on accented syllables. In the formulation of the model, which is defined in the log *F0* domain, each new prosodic phrase is preceded by an impulsewise phrase command of magnitude *Ap*, which quantifies the amount of reset of the declination line, and each pitch accent is associated with a box-shaped accent command of amplitude *Aa* which directly relates to the time and F0 intervals of the peak pattern that spans the respective syllable. The phrase command onset time *T0* is typically related to the segmental onset of the ensuing prosodic phrase, while the accent command onset time *T1* and offset time *T2* can be related to the timing of the underlying segments, typically syllables.

These are Fujisaki model parameters which are displayed underneath the *F0* contour in Figure 2 and Figure 3. Hence each parameter set reflects *F0* realized in a certain domain, for instance, in one and the same word, and the solid line following the extracted F0 pattern represents the output of the Fujisaki model for the given utterance. As we can see, the model contour smoothly approximates the natural contour and interpolates through unvoiced sounds. The Fujisaki model parameters are estimated from the natural F0 contour using an automatic algorithm [35].

Table 1: *Mean, S.D., and N for Aa, accent command duration (Acd, s), Fb (Hz), Ap and word duration (s).*

speaker		*Aa*	Acd	*Fb*	*Ap*	word dur
SJ	mean	.50	.422	117.49	.47	.268
	S.D.	.24	.178	18.11	.21	.192
	N	1087	1087	101	577	2970
MZ	mean	.47	.423	95.13	.55	.222
	S.D.	.23	.185	14.31	.22	.158
	N	1112	1112	103	527	3645

In SJ's utterance, for instance, we see accent commands associated with the words "again", "taking", and "advantage", see Figure 2. Based on the Fujisaki model parameters we can perform a comparative analysis of speech corpora and see which significant difference arise between speakers or speaking styles. By viewing the *F0* contour as a superposition of utterance, phrase and accent level components, we are able determine whether "high pitch" as indicative of charismatic speech affects all components or only one of them. Furthermore, it becomes possible to re-synthesize F0 contours and examine whether differences are also captured perceptually.

The perceptual local speech rate (*plsr*) is a psychophysical measure which was developed by [36] because earlier measures such as the local syllable rate and the local phone rate are not well-correlated, meaning that they represent different aspects of speech rate. Perception experiments with short stretches of speech being judged on a rate scale revealed that neither syllable rate nor phone rate is sufficient to predict the perception results. Subsequently it was shown that a linear combination of the two measures yielded a correlation of r=0.91 and a mean deviation of 10% which is accurate enough to successfully extract *plsr* from large spoken language corpora. The result is a smooth contour of local speech rate values aligned with the speech signal where a value of 100% represents a typical average speech rate while 50% being approx. half of it and 200% being roughly twice the average. It was also shown [37] that the language background affects the perception of local speech rate. German and Japanese listeners overshoot the speech rate of the respective unknown language by 7.5% or 9.1%, respectively. This additional deviation seems to be small enough to apply the *plsr* extraction method also to English speech. In the context of the current study, we are able to see whether the speakers differ with regards to their timing properties, that is, their dynamic use of speech rate, by not only looking at their mean syllable rate, but also at their perceptual tempo variations.

We determined the perceptual local speech rate following [36], calculating the rate as a percentage: *plsr[%]=8.6 x syllable rate+3.6 x phone rate-0.2*. The syllable and phone rates were determined using a moving window of 625ms at a step of 10ms, ignoring silent pauses.

4. Results of Analysis

In order to relate the Fujisaki model parameters to the segmental structure of the underlying utterances we align them to the boundaries of words or syllables. Subsequently, phrase commands can be related to the beginnings of prosodic phrases and accent commands to accented syllables or phrase-final boundary tones. We first performed a statistical analysis of distributions for the base frequency *Fb*, phrase command magnitude *Ap* and accent command amplitude *Aa* depending on the speaker, being either SJ or MZ, or the audience, being either customers and investors, to see whether they are significantly different. Furthermore, we looked at accent command and word durations. We extracted a total of 1,087 accent commands for SJ and 1,112 accent commands for MZ, and 201 phrase commands for SJ and 177 phrase commands for MZ. Table 1 provides an overview of the results.

We can see a few clear differences between the two speakers. MZ's base frequency is much lower than SJ's, but his phrase command amplitudes are higher than SJ's. In contrast, SJ's accent command amplitudes are significantly larger (Kruskal-Wallis test, $p<.001$ for all differences). This suggests that SJ's high pitch is related to a higher F0 floor *and* larger excursions on the accent level, that is, larger pitch accent ranges, whereas MZ starts from a lower F0 floor, but resets the declination line at phrase onsets more strongly. However, the durations of accent commands underlying the pitch accents (mean ~420ms) are similar for the two speakers, regardless of the fact that the 3.7/s word rate of SJ is significantly lower than MZ's 4.5/s. That is, the F0 gestures are equally long for both speakers, but due to the mathematical formulation of the Fujisaki model, the higher accent command amplitude *Aa* for SJ also results in steeper F0 slopes (cp. the last two accent commands in Fig.3).

Table 2: *Mean and S.D. for Aa, accent command duration (Acd), Fb (Hz), and Ap by audience type.*

	audience		*Aa*	Acd	*Fb*	*Ap*
SJ	customers	mean	.53	.414	115.49	.50
		S.D.	.24	.178	18.10	.21
	investors	mean	.47	.417	120.18	.44
		S.D.	.24	.176	17.25	.20
MZ	customers	mean	.45	.413	93.09	.50
		S.D.	.23	.178	9.53	.20
	investors	mean	.49	.422	94.70	.60
		S.D.	.24	.191	12.56	.23

Now, we look at the variability of intonation gestures. If we examine the raw pairwise variability index [38] between consecutive accent commands, that is the average of $|Aa_i\text{-}Aa_{i+1}|$, it is slightly larger for SJ (mean=.234) than for MZ (mean= .219), but the difference only approaches significance (Mann-Whitney-U-test, p=.066). The pairwise variability index between consecutive phrase commands, that is, the average of $|Ap_i\text{-}Ap_{i+1}|$ is larger for MZ (mean=.212) than for SJ (mean= .186), and this difference significant (Mann-Whitney-U-test, p=.008). These results mirror the amplitude differences we found above, larger in the accent domain for SJ, and larger in the phrase domain for MZ. Regarding the differences of Fujisaki model parameters as a function of audience type, Table 2 shows that there are striking and significant differences between SJ's customer- and investor-oriented speech. They concern in *Fb*, *Ap* and *Aa*. When addressing customers SJ "pitches" phrases and accents at higher *Ap* and *Aa* levels, and by simultaneously lowering *Fb*, he also increases his pitch range. MZ, on the other hand, shows a diametrically opposed behavior. While keeping his *Fb* level constant, he boosts his phrase resets and Aa values significantly when talking to the investors in the audience (Mann Whitney-U-test, p<.024).

Furthermore, we looked at the distance between consecutive phrase commands for the two speakers and found that SJ produces - despite his slower speech rate - more phrase commands (mean/S.D. distance 2.05/0.86 seconds) than MZ (2.13/ 0.86 seconds). But, this difference is not significant (Mann-Whitney-U test, p=0.180). Examining the distance between consecutive accent commands we find that MZ produces slightly more accents (mean/S.D. distance 0.80/0.56 seconds) than SJ (0.83/0.59 seconds). This difference, however, is also not significant either (Mann-Whitney-U test, p=0.171).

Table 3 displays means and S.D. of perceived local speech rate, as well as its absolute delta values, that is, the unsigned difference between adjacent 625ms windows of measurement spaced at 10ms intervals, along with syllable and phone rates, as a function of audience type. It is striking that the total mean of measurements for the two speakers is 99.1% with SJ's speaker mean being at 90.2% and MZ's at 107.7%. This suggests that the MZ is a faster-than-average speaker, whereas SJ speaks slower one than average. MZ also accelerates and decelerates significantly faster than SJ as is indicated by the *absolute* Δ*plsr*. Mann-U-tests of independent samples con-firm that all these differences are significant (p<0.01).

Table 3: *Mean and S.D. for perceived local speech rate (plsr [%]), its absolute delta [%], syllables/s, and phone/s, broken down for SJ/MZ by audience type.*

	audience		plsr	syl/s	phon/s	|Δplsr|
SJ	customers	mean	91.98	5.15	13.30	1.07
		S.D.	34.58	2.22	4.98	1.23
	investors	mean	89.03	4.96	12.95	1.04
		S.D.	36.43	2.38	5.15	1.22
MZ	customers	mean	108.60	6.09	15.68	1.22
		S.D.	36.09	2.29	5.14	1.31
	investors	mean	106.95	5.97	15.50	1.21
		S.D.	39.16	2.54	5.37	1.29

Both speakers speak significantly slower in their investor talk than in their customer talk (Mann-Whitney-U test, p<0.01). However, unlike MZ, SJ adjusts the dynamics of his speech rate, creating a higher absolute *Δplsr* when talking to customers. This seems plausible, as one way of creating tension and excitement in a discourse is by accelerating and decelerating depending on the importance of the current piece of information. We finally examined the perceived local speech rate con-tours. To this end we segmented the contours into pieces of either rising or falling slopes. A new segment starts when the gradient (i.e., Δ*plsr*) changes polarity. Hence, positive polarity indicates acceleration of speech rate, and negative polarity deceleration. For SJ, the average accelerating segment lasts for 198ms and decelerating one for 211ms. As *plsr* is calculated at 10ms intervals, the mean Δ*plsr* values of SJ are +0.89%/ 10ms and -0.88%/10ms, respectively. The corresponding values of MZ are +0.95%/10ms and -1.01%/10ms. Thus, accelerating segments of MZ last for 163ms and decelerating ones for 197ms on average. That is, SJ employs speech rate changes more symmetrically, balancing acceleration and deceleration phases, whereas MZ produces longer and steeper decelerating segments. Except for the slope in acceleration phases (Mann-Whitney-U test, p=0.229), SJ and MZ are significantly different in this respect (Mann-Whitney-U test, p<0.01).

5. Discussion and Conclusions

Charismatic speech is an important pillar of success and satisfaction in personal life and professional life - and it is a learnable competence [5,7,10]. However, in order to properly assess speakers' skills and deficits and effectively train and improve their charismatic speech, researchers and coaches need detailed insights into what charismatic speech actually means in phonetic or prosodic terms, and which characteristics need to be varied and adapted to what contexts [12]. Previous studies have primarily been concerned with identifying basic charisma-relevant prosodic parameters and with determining how general level shifts in these basic parameters affect charisma. For example, it was found that a higher F0 level, a larger F0 range, a faster speech rate, as well as a generally higher level of variability are beneficial for perceiving a speaker as charismatic [13,14,15,16,17,18].

The present study builds upon these findings. We modeled the behavior of a more charismatic speaker (SJ) in comparison to a less charismatic speaker (MZ) in terms of the prosodic Fujisaki parameters. By means of this contrastive quantitative modeling of local prosodic characteristics, our goal was to obtain indications of which strategies are in detail behind the charisma-triggering parameter level shifts. These indications can then be further examined in controlled production and perception experiments for their effectiveness and learnability.
Our results allow initial answers to three questions.

(1) What strategy can underlie a general increase in F0 level? SJ's data shows that a higher F0 level does not necessarily mean a general and with increasing values disproportionately stronger up-scaling of all F0 values. On the contrary, it can mean lowering the F0 baseline and instead modifying the F0 slopes of pitch accents and initial and final boundary tones such that they get longer, higher, and arrive faster at a high F0 level (see also the note on SJ's convexly shaped F0 contours in [39]). This strategy, applied by SJ, effectively combines a higher F0 level with a larger F0 range.

(2) Which sources of variability are particularly important for charismatic speech? SJ's data suggest that a strong and symmetric acceleration and deceleration in speech rate as well as high diversity in the alignment and scaling of pitch accents are particularly important. The latter strategy makes sense also because it is associated with phonological changes, whereas the variation in the scaling of phrase-initial F0 movements realized by MZ is phonologically irrelevant and hardly suitable to counteract an intonational monotony. In this context, our results could point to the central flaw of MZ that makes him sound less charismatic than SJ: MZ seems to focus too much on prosodic-boundary and topic signals and the syntagmatic structuring of his speech, while neglecting the expressive-emotional features that are crucial to charisma.

(3) Is audience type a relevant context factor? Yes, the strongly audience-sensitive prosodic behavior of both speakers and SJ in particular, is a strong indication that charismatic speaking could mean addressing different audience groups differently. Whether these differences are qualitative or quantitative in nature (as is suggested by our data) is another important question to be addressed in follow-up studies. The fact that SJ and MZ put more effort into sounding charismatic for customers or investors, respectively, fits in well with the two speakers' different business backgrounds and socializations.

6. References

[1] Pentland A., *Honest Signals - How they shape our world*. Cambridge: MIT Press, 2008.

[2] Towler A.J., “Effects of charismatic influence training on attitudes, behavior, and performance,” *Personnel Psychology 56*, pp. 363-381, 2003.

[3] Davis B.C, Hmieleski K.M, Webb J.W, Coombs J.E., "Funders' positive affective reactions to entrepreneurs' crowd-funding pitches: The influence of perceived product creativity and entrepreneurial passion," *J. of Bus. Vent. 32*, pp. 90–106, 2017.

[4] Gélinas-Chebat C., Chebat J.C., Vaninsky A., “Voice and Ad-vertising : Effects of Intonation and Intensity of Voice on Source Credibility, Attitudes and the Intend to Buy,” *Perceptual and Motor Skills 83*, pp. 243-262, 1996.

[5] Fox Cabane, O., *The Charisma Myth - How anyone can master the art and science of personal magnetism*. New York: Penguin.

[6] Jacobs, E., "Can you learn how to be charismatic?" *The Financial Times*, 2017.

[7] Antonakis J., Fenley M., Liechti S., “Can charisma be taught? Tests of two interventions,” *Acad. Manag. Learn. Educ. 10*, pp. 374–396, 2011.

[8] Potts, J., *A history of charisma*. London: Palgrave MacMillan, 2009.

[9] Awamleh R., Gardner, W.L., "Perceptions of leader charisma and effectiveness: The effects of vision content, delivery, and organizational performance. *The Leadership Quarterly 10*, pp. 345–373, 1999.

[10] Antonakis J., Bastardoz N., and Jacquart P., "Charisma: An ill-defined and ill-measured gift,” *Annual Review of Organization-al Psychology & Organizational Behavior 3*, pp. 293-319, 2016.

[11] Rosenberg A. and Hirschberg J., “Acoustic/Prosodic and lexical correlates of charismatic speech,” *Proc. Eurospeech 2005, Lisbon, Portugal*, pp. 1–4, 2005.

[12] Niebuhr O., Tegtmeier S., Brem A., “Advancing research and practice in entrepreneurship through speech analysis – from descriptive rhetorical terms to phonetically informed acoustic charisma metrics,” *J. of Speech Sciences 6*, pp. 3–26, 2017.

[13] Touati P., “Prosodic aspects of political rhetoric,“ *Proc. ESCA Workshop on Prosody, Lund, Sweden*, pp. 168–171, 1993.

[14] Biadsy F., Rosenberg A., Carlson R., Hirschberg J., Strangert E., “A cross-cultural comparison of American, Palestinian, and Swedish perception of charismatic speech,” *Proc. Speech Prosody, Campinas, Brazil*, pp. 579–582, 2008.

[15] Signorello R., D’Errico F., Poggi I., and Demolin D., "How charisma is perceived from speech. A multidimensional approach," *Proc. ASE/IEEE international conference on social computing, Amsterdam, Netherlands*, pp. 435-440, 2012.

[16] D'Errico F., Signorello R., Demolin D., Poggi I, “The perception of charisma from voice. A crosscultural Study,” *Proc. Humaine Association Conference on Affective Computing and Intelligent Interaction, Geneva, Switzerland*, pp. 552–557, 2013.

[17] Rosenberg A., Hirschberg J., ”Charisma perception from text and speech,” *Speech Communication 51*, pp. 640–655, 2009.

[18] Hiroyuki T., Rathcke T., ”Then, What is Charisma? The Role of Audio-visual Prosody in L1 and L2 Political Speeches,” *Proc. Phonetik & Phonologie, Munich, Germany*, pp. 1–3, 2016

[19] Bosker H.R., “The role of temporal amplitude modulations in the political arena: Hillary Clinton vs. Donald Trump,” *Proc. 18th Interspeech, Stockholm, Sweden*, pp.1–5, 2017.

[20] Niebuhr O., "On the phonetics of intensifying emphasis," *Phonetica 67*, pp. 170-198, 2010.

[21] K.W. Mortensen, *The Laws of Charisma: How to Captivate, Inspire, and Influence for Maximum Success*. New York: Amacom, 2011.

[22] R. Camper Bull, *Moving from Project Management to Project Leadership: A practical guide to leading groups*. Boca Raton: CRC, 2010.

[23] Weber, M., *The Theory of Social and Economic Organization*. New York: Free Press, 1947.

[24] Holladay S.J., Coombs W.T., “Communicating visions: An exploration of the role of delivery in the creation of leader charisma,” *Man. Com. Quart. 6*, pp. 405-427, 1993.

[25] Berger S., Niebuhr O., Peters B., “Winning Over an Audience – A Perception-based Analysis of Prosodic Features of Charisma-tic Speech,” *Proc. 43rd Annual Conference of the German Acoustical Society, Kiel, Germany*, pp. 1454–1457, 2017.

[26] Novák-Tót E., Niebuhr O., Chen A., ”A gender bias in the acoustic-melodic features of charismatic speech?” *Proc. 18th Interspeech, Stockholm, Sweden*, pp. 2248–2252, 2017.

[27] Gussenhoven, C., "Intonation and interpretation: Phonetics and phonology," Proc. 1st International Conference of Speech Prosody, Aix-en-Provence, France, pp. 43-53, 2002.

[28] Ohala, J.J., "An ethological perspective on common cross-language utilization of F0 in voice," *Phonetica 41*, pp. 1-16, 1984.

[29] P. Boersma, “Praat, a system for doing phonetics by computer”, *Glot International 5*, 341-345, 2001.

[30] Conger J.A., Kanungo R.N., *Charismatic Leadership in Organizations*. London: SAGE, 1998.

[31] Meyer, M., *What is Rhetoric?* Oxford: OUP, 1989.

[32] Strunk J., Schiel F., Seifart, F., "Untrained forced alignment of transcriptions and audio for language documentation corpora using WebMaus," *Proc. 9th International Conference on Language Resources and Evaluation, Reykjavik, Iceland*, pp. 3940-3947, 2014.

[33] Reddy S., Stanford J., "Toward completely automated vowel extraction: Introducing DARLA," Linguistics Vanguard, 2015.

[34] H. Fujisaki, K. Hirose, “Analysis of voice fundamental fre-quency contours for declarative sentences of Japanese”, Journal of the Acoustical Society of Japan 5, 233-241, 1984.

[35] H. Mixdorff, “A new approach to the fully automatic extraction of Fujisaki model parameters”, *Proceedings of ICASSP 2000*, vol. 3, 1281-1284, Istanbul, Turkey, 2000.

[36] Pfitzinger, H.R., "Local Speech Rate Perception in German Speech," *Proc. 14th ICPhS, San Francisco, USA*, pp. 893-896, 1999.

[37] Pfitzinger, H.R., Tamashima, M., "Comparing Perceptual Local Speech Rate of German and Japanese Speech," *Proc. of the 3rd Int. Conf. on Speech Prosody, Dresden, Germany*, pp. 105-108, 2006.

[38] Grabe, E., Low, E.L., "Durational variability in speech and the rhythm class hypothesis," in Gussenhoven, C., Warner, N. (eds.), *Papers in Laboratory Phonology 7*, Berlin: Mouton de Gruyter, pp. 515-546, 2012.

[39] Niebuhr, O., Thumm J., Michalsky J., "Shapes and timing in charismatic speech - Evidence from sounds and melodies," *Proc. 9th International Conference of Speech Prosody, Poznan, Poland*, submitted.